



Document Noise Removal using Sparse Representations over Learned Dictionary

Thanh Ha Do, Salvatore Tabbone, Oriol Ramos Terrades

► To cite this version:

Thanh Ha Do, Salvatore Tabbone, Oriol Ramos Terrades. Document Noise Removal using Sparse Representations over Learned Dictionary. The 13th ACM Symposium on Document Engineering, Sep 2013, Florence, Italy. 10.1145/2494266.2494281 . hal-00939174

HAL Id: hal-00939174

<https://inria.hal.science/hal-00939174>

Submitted on 13 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Document Noise Removal using Sparse Representations over Learned Dictionary

Thanh-Ha Do
Université de Lorraine - LORIA
- UMR 7503
Campus scientifique - BP 239
Vandoeuvre-lès-Nancy, France
ha-thanh.do@loria.fr

Salvatore Tabbone
Université de Lorraine - LORIA
- UMR 7503
Campus scientifique - BP 239
Vandoeuvre-lès-Nancy, France
tabbone@loria.fr

Oriol Ramos Terrades
Computer Vision Centre
08193 Bellaterra (Cerdanyola)
Barcelona, Espanya
oriolrt@cvc.uab.cat

ABSTRACT

In this paper, we propose an algorithm for denoising document images using sparse representations. Following a training set, this algorithm is able to learn the main document characteristics and also, the kind of noise included into the documents. In this perspective, we propose to model the noise energy based on the normalized cross-correlation between pairs of noisy and non-noisy documents. Experimental results on several datasets demonstrate the robustness of our method compared with the state-of-the-art.

Categories and Subject Descriptors

J.m [Computer Applications]: Miscellaneous

Keywords

Sparse Representation, Learned Dictionary, K-SVD, Normalized Cross Correlation, Noise Suppression

1. INTRODUCTION

The performance of many pattern recognition techniques applied on images depends on an accurate control of the noise included into the image. In the case of document image analysis applications, the kind of noise is different compared to the noise found in natural scenes images that are generated by devices like digital cameras or similar.

Therefore, the problem of document denoising has been tackled since the very beginning. There is a vast literature on methods proposing solutions such as median filter [6], morphological filters [16], or curvelets transform [18]. Median filter replaces each pixel in the noisy image by the median of pixels in a neighborhood of that pixel, while morphological filter carries out dilation and erosion operations as denoisy operators. In addition, morphological filtering can discriminate between positive and negative noise spikes, whereas median filter cannot [16]. Both filters, median and

morphological, are appealing because they are easy to implement and perform well with the presence of impulse noise. However, neither of them are efficient for other types of noise like white noise or noise arising from printing, photocopying and scanning processes. This kind of noise not only generally causes undesirable document appearance, but also has a bad influence on document processing performance.

Recent researches obtain good performance in denoising gray-scale images with white noise using Multi Resolution Analysis (MRA) methods [18]. Sparse transforms and MRA methods are applied to a wide range of image processing problems as image compression, image restoration and image denoising [18]. These methods have proven to perform well in terms of Mean Square Error (MSE) measure as well as Peak Signal-to-Noise Ratio (PSNR) measure for essentially white noise (additive and following a Gauss distribution). Moreover, sparse transforms, like curvelets, contourlets, wedgelets, bandelets, or steerable wavelets, have also been successfully applied in document images for removing noisy edges, showing that sparse representation can effectively be used for denoising purposes. Sparse transforms represent images as linear combinations of (atom) functions of a given particular family of functions (dictionary). However, the overall performance of these methods depends on two factors: first of all, a *a priori* knowledge about images which drives the choice of dictionary functions and secondly, the kind of noise found in document images. Recently, curvelets transform has been applied in document denoising with a relative high degree of success [11]. In that approach, the authors take advantage of directional properties of curvelets transform to denoise degraded graphic documents. The results obtained in that work show an improvement in removing noise for document images comparing with other state-of-the-art methods. However, curvelet is one of the pre-defined dictionaries, and therefore can only work well with some kinds of noise and cannot be adapted to arbitrary noise models.

In this paper, we address the task of document denoising by proposing a method which overcomes the difficulties found in denoising methods based on MRA. On the one hand, we apply the K-SVD algorithm to learn a proper set of atom functions adapted to both document characteristics and document noise. On the other hand we propose an energy noise model which allows us to easier set the threshold required for noise removal even if the noise model is unknown.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

DocEng '13, September 10–13, 2013, Florence, Italy.

Copyright 2013 ACM 978-1-4503-1789-4/13/09 ...\$15.00.

<http://dx.doi.org/10.1145/2494266.2494281>.

The remainder of this paper is organized as follows. We introduce the theoretical framework of sparse representations in section 2 and the learning algorithm in section 3. Then, we recall the document degradation models used in experiments (section 4) and we propose the energy noise model in section 5. Next, we discuss the experimental results in section 6 and finally, we give our conclusions in section 7.

2. SPARSE REPRESENTATION

The main idea of the proposed method is to find a dictionary adapted to the properties of the data which will allow us to obtain a denoised version of the original degraded images using sparse representations.

A sparse representation is a linear combination of few atoms (basis functions) of a given dictionary. Mathematically, given a dictionary A and a signal h , we consider the under-determined linear system of equations $h = Ax$, with $A = \{a_1, a_2, \dots, a_m\} \in R^{n \times m}$, $h \in R^n$, $x \in R^m$, $m \gg n$. If A is a full-rank matrix, there will be infinitely many different sets of values for the x_i 's that satisfy all equations simultaneously. The set of x can be described using mathematical language. However, from an application point of view, one of the main tasks in dealing with the above system of equations is to find the proper x that can explain h well comparing with others. To gain this well-defined solution, a function $f(x)$ is added to assess the desirability of a would-be solution x , with smaller values being preferred:

$$(P_f) : \min_x f(x) \text{ subject to } Ax = h \quad (1)$$

If $f(x)$ is the l_0 pseudo-norm $\|x\|_0$ (number nonzero elements in vector x), then the problem (P_f) becomes finding the sparse representation x of h satisfying:

$$(P_0) : \min_x \|x\|_0 \text{ subject to } Ax = h \quad (2)$$

In general, solving equation (2) is often difficult (NP-hard problem) and one of the choices is to look for an approximate solution using greedy algorithms, such as *Matching Pursuit* (MP) [15], *Orthogonal MP* (OMP) [17], *Weak MP* [19] and *Thresholding* algorithm [8]. A greedy algorithms is an algorithms that follows the problem solving heuristic of making the locally optimal single term updates with the hope of finding a global optimum. In our case, the set of active columns started from empty is maintained and expanded by one additional column of A at each iteration. The chosen column is a column that maximally reduces the residual l_2 error in approximating h from the currently active columns. The residual l_2 error is evaluated after constructing an approximate including the new column; if it is bounded below a specified threshold, the algorithm terminates.

The other choice is to relax the l_0 -norm by replacing it with the l_p -norms for some $p \in (0, 1]$ or by smooth functions such as $\sum_i \log(1 + \alpha x_i^2)$, $\sum_i x_i^2 / (\alpha + x_i^2)$, or $\sum_i (1 - \exp(-\alpha x_i^2))$. The interesting algorithm of this family is the *FOcal Under-determined System Solver* (FOCUSS) [10]. In this algorithm, the l_p -norm (for some fixed $p \in (0, 1]$) is represented as a weighted l_2 -norm by using *Iterative Reweighed Least Squares* (IRLS) method [5].

Another popular strategy is to replace the l_0 -norm by the l_1 -norm proposed by Donoho *et al* [7]

$$(P_1) : \min_x \|W^{-1}x\|_1 \text{ subject to } Ax = h \quad (3)$$

The matrix W is a diagonal positive-definite matrix. A natural choice for each entry in W is $w(i, i) = 1/\|a_i\|_2^{-1}$. Let $\tilde{x} = W^{-1}x$, then equation (3) is re-formulated as

$$(P_1) : \min_{\tilde{x}} \|\tilde{x}\|_1 \text{ subject to } h = A\tilde{x} = \tilde{A}\tilde{x} \quad (4)$$

in which \tilde{A} is the normalized version of A . Equation (4) is the classic basis pursuit format, and the solution x can be found by de-normalizing \tilde{x} . Thus, (P_1) is usually used with a normalized matrix. The solution for (P_1) problem can be found by some existing numerical algorithms, such as *Basis Pursuit* by Linear Programming [3] or IRLS (for $p = 1$).

If there exists some appropriate conditions on A and x , like

$$\|x\|_0 \leq \frac{1}{2} \left(1 + \frac{1}{\max_{i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}} \right)$$

then Basic Pursuit as well as OMP give the unique solution of (4) and it is also the unique solution of (P_0) .

However, when signals are perturbed by noise it becomes more useful to relax the exact constraint: $h = Ax$ by using instead the quadratic penalty function $Q(x) = \|Ax - h\|_2^2 \leq \epsilon$, with $\epsilon \geq 0$ being the error tolerance. Therefore, an error-tolerant version of (P_0) is defined by:

$$(P_0^\epsilon) : \min_x \|x\|_0 \text{ subject to } \|Ax - h\|_2 \leq \epsilon \quad (5)$$

In (P_0^ϵ) the l_2 -norm is used for evaluation, and the error $Ax - h$ can be replaced by other options, as l_1 , l_2 , or l_∞ .

Observe that problem (P_0^ϵ) as defined above is useful for the given task. Assuming that signal h has noise e with finite energy $\|e\|_2^2 \leq \epsilon^2$, $h = Ax + e$, solving (P_0^ϵ) can help us to find the solution \hat{x} . Then, we can recover the unknown denoised signal \hat{h} as $\hat{h} = A\hat{x}$. Similarly, when relaxing l_0 -norm to an l_1 -norm, we get (P_1^ϵ) known in the literature as *basis pursuit denoising* (BPDN) [7]

$$(P_1^\epsilon) : \min_x \|x\|_1 \text{ subject to } \|Ax - h\|_2 \leq \epsilon \quad (6)$$

3. LEARNED METHODOLOGY FOR DICTIONARY AND K-SVD ALGORITHM

The optimal solution of the problem (P_1^ϵ) directly depends on the used dictionary A . Our working hypothesis is that if we are able to learn A from a training dataset, then A should be well adapted to the document characteristics and noise. So, in this section, we review learning algorithms, in general, used for constructing a dictionary A and in particular, the learning algorithms we have used: the K-SVD algorithm.

In a general learning methodology, a family l signals $\{h_j\}_{j=1}^l$ is considered as the training database. Our goal is to find a dictionary A in which each signal $h_j \in R^n$ has an optimally sparse approximation $\hat{h}_j \simeq Ax_j$ satisfying $\|\hat{h}_j - h_j\|^2 \leq \epsilon$, or finding:

$$\min_{A, x_j} \sum_{j=1}^l \|x_j\|_1 \text{ subject to } \|h_j - Ax_j\|_2 \leq \epsilon, \text{ for all } j = 1, \dots, l \quad (7)$$

This dictionary can be obtained by the learning process that iteratively adjusts A via two main stages: sparse coding stage and update dictionary stage. In the sparse coding

$$\|x\|_2 \triangleq (\sum_{i=1}^M |x_i|^2)^{1/2}, \text{ with } x \in R^M$$

stage, all sparse representations $X = \{x_j\}_{j=1}^l \in R^{m \times l}$ of $H = \{h_j\}_{j=1}^l \in R^{n \times l}$ are found by solving equation (6), on the condition that A is fixed. In the update dictionary stage, an updating rule is used to optimize the sparse representations of the training signals. In general, the way to update the dictionary is different from one learning algorithm to another. There are two well-known dictionary-learning algorithms, named *Method of Optimal Directions* (MOD) by Engan *et al* [9], and K-SVD by Aharon *et al* [1]. These two algorithms behave similarly with a small advantage to the K-SVD [8]. So, we chose the K-SVD algorithm in our paper to construct the learned dictionary.

In K-SVD algorithm [1], the updating rule is to make a modification on dictionary's columns. At this step, we handle to update sequentially each columns a_{j_0} of A such that the residual error (8) is minimized, where X and $\{a_1, \dots, a_{j_0-1}, a_{j_0+1}, \dots, a_m\}$ are fixed,

$$\begin{aligned} \|H - AX\|_F^2 &= \|H - \sum_{j=1}^m a_j x_j^T\|_F^2 \\ &= \|(H - \sum_{j \neq j_0} a_j x_j^T) - a_{j_0} x_{j_0}^T\|_F^2 \quad (8) \\ &= \|E_{j_0} - a_{j_0} x_{j_0}^T\|_F^2 \end{aligned}$$

In equation (8), $x_{j_0}^T \in R^l$ is the k -th row in X and the notation $\|\cdot\|_F$ stands for the Frobenius norm. Because X and all the columns of A are fixed excepted the column a_{j_0} , so $E_{j_0} = H - \sum_{j \neq j_0} a_j x_j^T$ is fixed. It means that the minimum error $\|H - AX\|_F^2$ depends only on the optimal a_{j_0} and $x_{j_0}^T$. This is the problem of approximating a matrix E_{j_0} with another matrix which has a rank 1 based on minimizing the *Frobenius* norm. The optimal solutions $\tilde{a}_{j_0}, \tilde{x}_{j_0}^T$ can be given by the *Singular Value Decomposition* (SVD) of E_{j_0} of rank r_1 , namely

$$\tilde{a}_{j_0} \tilde{x}_{j_0}^T = Q_1 \tilde{U} Q_2$$

where $Q_1 = \{q_1^1, \dots, q_n^1\}$, $U = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{r_1})$, $Q_2 = \{q_1^2, \dots, q_l^2\}$ is the SVD of E_{j_0} : $E_{j_0} = Q_1 U Q_2$; and \tilde{U} is the same matrix as U except that it contains only one singular values σ_1 (the other singular values are replaced by zero). This means $\tilde{a}_{j_0} = q_1^1$ and $\tilde{x}_{j_0}^T = \sigma_1 q_1^2$. However, the new vector $\tilde{x}_{j_0}^T$ is very likely to be filled, implying that we increase the number of non-zeros in the representation of X , or the condition about the sparsity of X can be broken.

This problem can be overcome as follows. Define the group of indexes where $x_{j_0}^T$ is nonzero:

$$\omega_{j_0} = \{i | 1 \leq i \leq l, x_{j_0}^T(i) \neq 0\}.$$

and a matrix $\Omega_{j_0} \in R^{l \times |\omega_{j_0}|}$ is defined $\Omega_{j_0}(\omega_{j_0}(i), i) = 1$ and zeros elsewhere. Let

1. $x_{j_0}^R = x_{j_0}^T \Omega_{j_0}$, $x_{j_0}^R \in R^{|\omega_{j_0}|}$
2. $E_{j_0}^R = E_{j_0} \Omega_{j_0}$, $E_{j_0}^R \in R^{n \times |\omega_{j_0}|}$

and following equation (8), the minimization is equivalent to

$$\|E_{j_0} \Omega_{j_0} - a_{j_0} x_{j_0}^T \Omega_{j_0}\|_F^2 = \|E_{j_0}^R - a_{j_0} x_{j_0}^R\|_F^2 \quad (9)$$

Note that the solution of (9) $\tilde{x}_{j_0}^R$ has the same support as the original $\tilde{x}_{j_0}^T$, and the optimal values $\tilde{x}_{j_0}^R, \tilde{a}_{j_0}$ can be obtained by finding SVD of a subset of the columns of the

error matrix $E_{j_0}^R$ of rank r_2 : $E_{j_0}^R = S D V^T$. The solution for \tilde{a}_{j_0} is defined as the first column of S , and the coefficient vector $\tilde{x}_{j_0}^R$ as the first column of V multiplied by d_1 , with $D = \text{diag}(d_1, d_2, \dots, d_{r_2})$. More details about the K-SVD algorithm can be found in algorithm 1.

Algorithm 1 Learning algorithm K-SVD

INPUT: $A_{(0)} \in R^{n \times m}$; $\{h_i\}_{i=1}^l$; $k = 0$;
1. Initialize: Normalization the columns of matrix $A_{(0)}$;
2. Main Iteration
 $k = k + 1$;
while ($\|H - A_{(k)} X_{(k)}\|_F^2$ is not small enough) **do**
 - Solve (P_1^ϵ) to find all sparse representation $\{x_i\}_{i=1}^l$ of $\{h_i\}_{i=1}^l$
 for ($j_0 = 1$ to m) **do**
 - Define: $\omega_{j_0} = \{i | 1 \leq i \leq l, x_{j_0}^T(i) \neq 0\}$
 - Calculate E_{j_0} via equation $E_{j_0} = H - \sum_{j \neq j_0} a_j x_j^T$
 - Let $E_{j_0}^R$ is the sub-matrix of E_{j_0} on the columns corresponding to ω_{j_0}
 - Calculate SVD of $E_{j_0}^R$: $E_{j_0}^R = S D V^T$.
 - Updating: $a_{j_0} = s_1$ and $x_{j_0}^R = d_1 v_1$ with $S = \{s_1, \dots, s_n\}$, $V = \{v_1, \dots, v_{|\omega_{j_0}|}\}$, $D = \text{diag}(d_1, d_2, \dots, d_{r_2})$
 end for
end while
OUTPUT: The result $A_{(k)}$

4. DOCUMENT DEGRADATION MODELS

Inspired by several authors like in [12, 11], we have used document noise models for evaluating the proposed method. In 1993, Kanungo *et al* [12] introduced a statistical model for document degradation showing quite realistic results, which gave the following three reasons justifying research on noise models. First of all, noise modeling allows to study recognition algorithms in general, as a function of the perturbation of the input data. Secondly, it permits the evaluation of any algorithm depending on the degradation level. Thirdly, a knowledge of the degradation model can enable us to design algorithms for image restoration. More recently, [2] has proposed the Noise Spread model, inspired on the physics of image acquisition process. According to this model the acquired image is obtained as a result of convolving the source image with the sensor function defined by the Point Spread Function with white noise.

Since the Kanungo noise model [12, 13] is a statistical model, widely used to verify the robustness of document image analysis methods to noise, we use it as noise model in this paper. However, the extension to another model is easy as our method is enabled to denoise documents for which the model of noise is unknown. One of the advantages of degradation models is they permit to generate degraded images controlled by the parameter models. The qualitative results of these images range from quite realistic noisy images to unrealistic, or even highly degraded images in function of the parameters set up. For instance, in Figure (1), symbol images (b) and (d) provide more realistic symbol degradation than (c), (e), (f) and (g).

Bi-level images are represented by white background pixels and black foreground pixels representing the different entities of the document. The Kanungo model needs six pa-

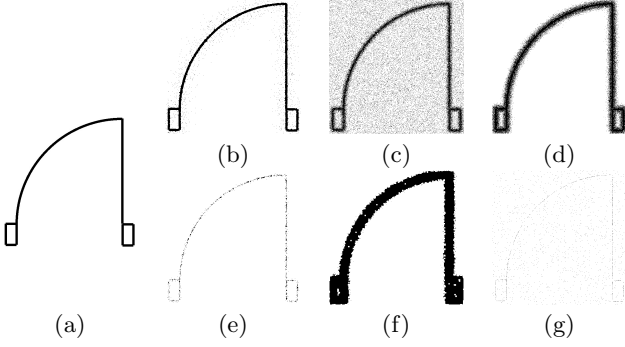


Figure 1: (a): Original binary symbol; from (b) to (g) examples of six levels of Kanungo noise of the GREC 2005 dataset.

parameters α_0 , α , β_0 , β , η , and k as a function of the pixel distance to the shape boundaries to degrade a binary image. α and α_0 control the probability of flipping a foreground pixel to a background pixel, in other words, these two parameters provide the probability of changing a black pixel to a white one. Similarly, parameters β and β_0 control the probability of changing a background (white) pixel with a foreground (black) pixel. In addition, these two probabilities exponentially decay on the distance of each pixel to the nearest boundary pixel. In contrast, the parameter η is a constant value added to all pixels regardless their relative position to shape boundaries. Finally, the last parameter k is the size of the disk used in the morphological closing operation. The whole process of image degradation can be summarized in the following three steps:

1. Use standard distance transform algorithms to calculate the distance d of each pixel from the nearest boundary pixel.
2. Each foreground pixel and background pixel is flipped with probability $p(0|1, d, \alpha_0, \alpha) = \alpha_0 e^{-\alpha d^2} + \eta$, and $p(1|0, d, \beta_0, \beta) = \beta_0 e^{-\beta d^2} + \eta$
3. Use a disk structuring element of diameter k to perform a morphological closing operation.

5. ENERGY NOISE MODEL

For denoising images using basis pursuit denoising, we need to decide which kind of dictionary A is used along with the best value ϵ in equation (6). From Section 3 we know how to learn a dictionary A adapted to noisy data. In this Section, we explain how to choose the best value ϵ when we apply equation (6) on the image patches with size w .

Denoising in MRA methods assumes that images have been corrupted by an additive white noise. For such noisy images, the energy of noise η is proportional to both the noise variance and size image. For Noise Spread model, authors in [11] empirically supposed that the optimal energy of noise depends on the noise spread relation. However, neither of these two assumptions can be applied to document images where the noise follows a Kanungo noise model instead of white noise. The reason is that in Kanungo model pixels near the shape boundaries have higher probability to

be affected by noise than pixels far from the shape boundaries. On the contrary, white noise model assumes statistical independence between the noise and the image. In fact, the probability that a pixel is perturbed by noise depends only on the variance of the model and does not depend on the position of a pixel in the image. In addition, the parameters used in Noised Spread model are different from the Kanungo's ones, so we cannot use the noise spread relation proposed in [11] to decide the value for ϵ .

Therefore, we propose an energy noise model inspired by [14]. Thus, we evaluate the noise level using the peak values of the normalized cross-correlation between noisy and cleaned documents. Let D being a training dataset including $2t$ documents (D_i^c, D_i^n) , $i = 1, \dots, t$ where D_i^c is a cleaned document and D_i^n is its noisy version. We define r_i as the peaks of normalized cross-correlation between D_i^c and D_i^n , then the tolerance error value is defined by:

$$\epsilon = cw\bar{r} \quad (10)$$

where c is a constant value set experimentally, w is the size of patches, \bar{r} is the mean value of the peaks r_i .

We can summarize the procedure for document denoising using sparse representation of a learned dictionary A as follows:

1. Create a training database using a sliding window of size $w \times w$ to scan the corrupted image $y \in R^{M \times N}$ with scanning step set to 1 pixel in both directions. All $(M - w + 1)(N - w + 1)$ obtained patches $\{h_j\}_{j=1}^l$, $h_j \in R^{w \times w}$ are considered as the training database.
2. Create a learned dictionary using the K-SVD algorithm to create the learned dictionary A from $\{h_j\}_{j=1}^l$.
3. Combine the learned dictionary A with the sparse representation model in the purpose of denoising image, following:
 - a. Find the solution of the optimization problem (6) for each patch h_j

$$\hat{x}_j = \arg \min \|x_j\|_1 \text{ subject to } \|Ax_j - h_j\|_2 \leq \epsilon,$$
 - b. Compute the denoised version of each patch h_j by $\hat{h}_j = A\hat{x}_j$,
 - c. Merge the denoised patches \hat{h}_j to get the denoised image \hat{y} .
4. Binarise \hat{y} to get the final result \tilde{y} .

6. EXPERIMENTAL RESULTS

We firstly evaluated our algorithm on the GREC 2005 dataset. This dataset has 150 different symbols which have been degraded using the Kanungo's method to simulate the noise introduced by the scanning process. Six sets of parameters are used to obtain six different noise levels as shown in Figure (1).

At each level of noise, a dataset containing 2×50 noisy and cleaned symbols are used to calculate the value of \bar{r} and ϵ (Section 5). We empirically found that the best results in denoising bilevel images are achieved when c in equation (10) belongs to $[0.4, 1]$.

Figure (2) shows one example about the normalized cross-correlation of two images (a) and (b) with its maximum

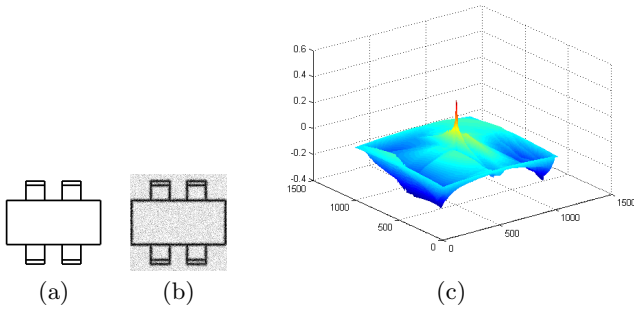


Figure 2: Normalized cross-correlation between two images.

| Level | 1 | 2 | 3 | 4 | 5 | 6 |
|-----------|--------|--------|--------|--------|--------|--------|
| \bar{r} | 0.9133 | 0.4412 | 0.5629 | 0.3698 | 0.4413 | 0.2006 |

Table 1: The value of \bar{r} at six level of noise.

value $r_i = 0.4106$; and table (1) presents the best values of \bar{r} corresponding to each level of degradation.

The learning dictionary was produced using K-SVD algorithm with 50 iterations, and the training dataset includes all 8×8 patches. Those patches were taken from a corrupted image h . The ratio of the dictionary is $1/4$ ($m = 4 \times n$).

To examine the effectiveness of the proposed method, we compare it with three existing methods for denoising binary images: median filtering, morphological operators (opening and closing), and curvelet transform. The median filtering is performed with a window size of 3×3 . The morphological operators use a 3×3 structuring element. Curvelets transform has been verified upon the best value of $\eta = c \times \sqrt{MN} \times \sigma^2$, $\sigma \in [0.02, 0.1]$ for each noise level, where M, N is the size of the image. The criterion to choose these best values σ is the average MSE (Mean Squared Error).

Moreover, using the traditional MSE to estimate the quality change between the original document and the reconstructed one, all algorithms are evaluated by Jaccard's similarity measure [4]. This measure is computed based on the three values a, b, c as below:

$$S = \frac{a}{a + b + c} \quad (11)$$

where

$$a = |\{(i, j) | y^0(i, j) = 1, \tilde{y}(i, j) = 1, 1 \leq i \leq M, 1 \leq j \leq N\}|$$

$$b = |\{(i, j) | y^0(i, j) = 0, \tilde{y}(i, j) = 1, 1 \leq i \leq M, 1 \leq j \leq N\}|$$

$$c = |\{(i, j) | y^0(i, j) = 1, \tilde{y}(i, j) = 0, 1 \leq i \leq M, 1 \leq j \leq N\}|$$

a means 'right matches', and b, c mean 'mismatches'; $y^0, \tilde{y} \in R^{M \times N}$ are cleaned and denoised images, respectively. The maximal value of the Jaccard measure is one when two images are identical.

Figure (3) shows an example of denoised images and table (2) and (3) show the average results obtained by the four methods on 6 levels of noise that are respectively evaluated by MSE and Jaccard's measures. A paired Wilcoxon signed test with a significance level of 5% is used also to check whether the difference between the results obtained by our

| | Median | OC | Curvelet | Proposed method |
|---------|-----------|-----------|------------------|-----------------|
| Level 1 | 0.926 (-) | 0.937 (-) | 0.963 (=) | 0.963 |
| Level 2 | 0.630 (-) | 0.177 (-) | 0.565 (-) | 0.655 |
| Level 3 | 0.428 (-) | 0.787 (-) | 0.389 (-) | 0.826 |
| Level 4 | 0.088 (-) | 0.001 (-) | 0.459 (+) | 0.135 |
| Level 5 | 0.261 (-) | 0.268 (-) | 0.263 (-) | 0.284 |
| Level 6 | 0.001 (-) | 0.000 (-) | 0.059 (=) | 0.060 |

Table 2: Average value gained by Jaccard's similarity measure.

| | Median | OC | Curvelet | Proposed method |
|---------|-------------|-------------|-------------------|-----------------|
| Level 1 | 2.896 (-) | 2.454 (-) | 1.403 (=) | 1.423 |
| Level 2 | 21.015(-) | 30.865 (-) | 28.421 (-) | 19.687 |
| Level 3 | 53.690 (-) | 8.899 (-) | 57.771 (-) | 8.008 |
| Level 4 | 33.599 (-) | 36.819 (-) | 19.815 (+) | 31.881 |
| Level 5 | 111.632 (-) | 107.358 (-) | 110.286 (-) | 98.212 |
| Level 6 | 37.863 (-) | 37.915 (-) | 35.632 (=) | 35.772 |

Table 3: Average values gained by MSE.

method and the ones obtained by the other methods is significant. In these tables, an entry mark by $(-)$ indicates that the corresponding method performs worst than our method. Similarly, an entry marked by $(+)$ indicates that the corresponding method outperforms the proposed method, and an entry marked by $(=)$ indicates that results obtained by the both methods are not significantly different.

Table (2) and (3) also show that at level 5 and level 6, none of the four methods are good enough but other methods are worse than ours. We further examine the set of noisy images at level 4 and we found that the set of noisy patches of the corrupted image cannot provide a good training data since most of patches are trivial (zeros value), making not enough discrimination between $A_{(0)}$ and $A_{(k)}$ (see algorithm 1). This can explain why the curvelets transform method is better than the proposed method at this level of noise. Although at level 1 the Wilcoxon signed test indicates that results obtained by the curvelets and our method are not significantly different, when we zoom the denoised images we found that the intersection of edges are not well restored with curvelet as shown in Figure (4). Since curvelets are smooth functions, they are not well adapted for singularity points.

We also verified our method on real scanned documents created by printing the original documents and scanning printed documents at the different resolutions. By using 15 original documents, we got 12 scanned datasets with different resolutions. Figure (6) shows an example of one of the scanned documents.

The aim is to test our method on real documents with an unknown model of noise twelve scanned images are tested and evaluated using the Structural similarity (SSIM) index. SSIM [20] is also a method for measuring the similarity that is designed to improve MSE, which have proved to be inconsistent with human eye perception. SSIM metric is computed on various windows of images. The distance between two windows s_i and q_i is calculated following the equation (12) where s_i, q_i are taken from the same location of the noisy image and reconstructed image.

$$S(s_i, q_i) = \left(\frac{2\mu_{s_i}\mu_{q_i} + c_1}{\mu_{s_i}^2 + \mu_{q_i}^2 + c_1} \right) \left(\frac{2\sigma_{s_i}\sigma_{q_i} + c_2}{\sigma_{s_i}^2 + \sigma_{q_i}^2 + c_2} \right) \quad (12)$$

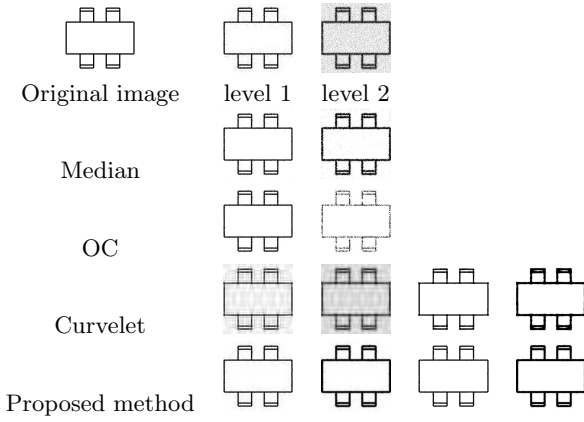


Figure 3: Results of denoising the noisy images with Kanungo model at levels 1 and 2 of degradation. Columns 2 and 3 are the denoised images following each method. Columns 4 and 5 are the binarized denoised images of columns 2 and 3, respectively. For the median and OC, images are already binarized in columns 2 and 3.

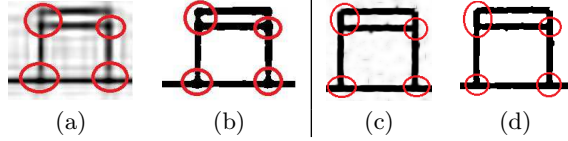


Figure 4: (a), (c): Zoom of denoised images by curvelets and our method, respectively. (b), (d) denoised binary version respectively of (a) and (c).

$\mu_{s_i}, \mu_{q_i}, \sigma_{s_i}, \sigma_{q_i}$ are the average local and the sample standard deviations of s_i and q_i , respectively; $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ are two variables to stabilize the division with weak denominator; L is the dynamic range of the pixel-values; and k_1, k_2 are two constants. The resultant SSIM index is a decimal with a value 1 in the case of two identical sets of data. In this paper, SSIM index is calculated on a window size 8×8 , and the values of L, k_1, k_2 are respectively 100, 0.01 and 0.03. Table (4) presents the results in comparison with other methods. We can see that in each case the performance of our approach is good compared to the other.

The last experiment is done on the DIBCO 2009 dataset. This dataset contains images that range from grayscale to color and from real to synthetic. The value of \bar{r} for this experiment equals 0.7321 and is calculated as the same way as described above. Figure (6) presents one learning dictionary build by corrupted patches with size 8×8 on DIBCO images. Figure (7a) gives the documents in DIBCO dataset and its denoised versions got by our approach (Figures (7b) and (7c)).

As the experiment before, the SSIM measure is used with the purpose of comparison. Table (5) presents the results for the 5 handwritten images shown in Figure 7 and we can see that in each case our approach provides a good result. Tables (4) and (5) present the difference results (the last rows) using also a paired Wilcoxon signed test with a significance level of 5%. We can observe that the difference between our approach performance compared to the other methods per-

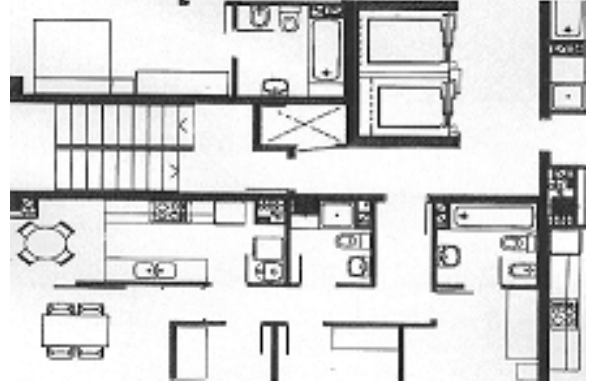


Figure 5: An example of a scanned documents.

| Images | Median | OC | Curvelet | Proposed method |
|---------|------------|------------|------------|-----------------|
| 1 | 0.5374 | 0.4640 | 0.4693 | 0.6834 |
| 2 | 0.6906 | 0.5165 | 0.6184 | 0.7534 |
| 3 | 0.5749 | 0.5158 | 0.5057 | 0.7421 |
| 4 | 0.6482 | 0.4879 | 0.5666 | 0.7454 |
| 5 | 0.6008 | 0.5056 | 0.5186 | 0.7458 |
| 6 | 0.6310 | 0.4409 | 0.5320 | 0.7145 |
| 7 | 0.6018 | 0.5030 | 0.5244 | 0.7580 |
| 8 | 0.6066 | 0.4814 | 0.5199 | 0.7381 |
| 9 | 0.6487 | 0.4662 | 0.5638 | 0.7456 |
| 10 | 0.6203 | 0.4603 | 0.5377 | 0.7317 |
| 11 | 0.6946 | 0.4832 | 0.5987 | 0.7685 |
| 12 | 0.7264 | 0.4494 | 0.6274 | 0.7733 |
| Average | 0.6320 (-) | 0.4812 (-) | 0.5485 (-) | 0.7416 |

Table 4: The obtained results when comparing proposed method with structural similarity SSIM index.

formance is statistically significant even in the case where the noise model is unknown.

| Images | Median | OC | Curvelet | Ours approach |
|---------|------------|------------|------------|---------------|
| 1 | 0.6420 | 0.6926 | 0.7054 | 0.9528 |
| 2 | 0.4628 | 0.5155 | 0.5156 | 0.9784 |
| 3 | 0.5115 | 0.5315 | 0.5064 | 0.8648 |
| 4 | 0.4595 | 0.4946 | 0.4692 | 0.8933 |
| 5 | 0.6953 | 0.7283 | 0.7441 | 0.9416 |
| Average | 0.5542 (-) | 0.5925 (-) | 0.5881 (-) | 0.9261 |

Table 5: The obtained results with DIBCO 2009 dataset using SSIM measure.

To conclude this section, all evaluation measures have indicated that our method performs better than all of the other methods in most of the cases.

7. CONCLUSIONS

A novel algorithm for denoising document images by using learning dictionary based on sparse representation has been presented in this paper. Learning method starts by building a training database from corrupted images, and constructing an empirically learned dictionary by using sparse represen-

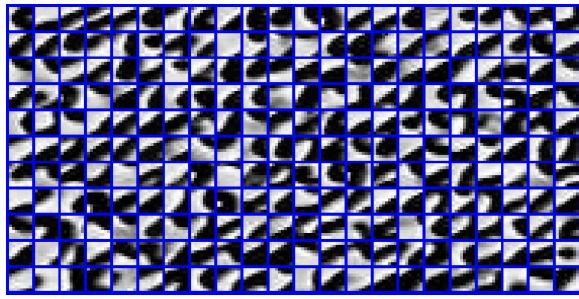


Figure 6: The trained dictionary on DIBCO images with $w = 8$.

tation. This dictionary can be used as a fixed dictionary to find the solution of the basis pursuit denoising problem. In addition, we provide a way to define the best value of the tolerance error (ϵ) based on a measure of fidelity between two images. The efficiency of ϵ has been also approved experimentally on different datasets for different resolutions and different kinds of noise. All experimental results show that our method outperforms existing ones in most of the cases.

8. ACKNOWLEDGEMENTS

Oriol Ramos Terrades, has been partially supported by the Spanish project TIN2012-37475-C02-02.

9. REFERENCES

- [1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.
- [2] E. Barney. Modeling image degradations for improving ocr. In *European Conference on Signal Processing*, pages 1–5, 2008.
- [3] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.
- [4] S. S. Choi, S. H. Cha, and C. Tappert. A survey of binary similarity and distance measures. *Journal on Systemics, Cybernetics and Informatics*, 8(1):43–48, 2010.
- [5] I. Daubechies, R. Devore, M. Fornasier, and C.S Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, october 2009.
- [6] E. Davies. *Machine Vision: Theory, Algorithms and Practicalities*. Academic Press, 1990.
- [7] D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proceeding of the National Academy of Sciences of the United States of America*, 100(5):2197–2202, 2003.
- [8] M. Elad. *Sparse and redundant representation: From theory to applications in signal and images processing*. Springer, Reading, Massachusetts, 2010.
- [9] K. Engan, S. O. Aase, and J. H. Husoy. Frame based signal compression using method of optimal directions (mod). In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1999.
- [10] I. Gonzalez and B. Rao. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *Signal Processing*, 45(3):600–616, March 1997.
- [11] V-T. Hoang, E.H. Barney Smith, and S. Tabbone. Edge noise removal in bilevel graphical document images using sparse representation. In *IEEE international conference on Image Processing*, 2011.
- [12] T. Kanungo, R. M. Haralick, and I. T. Phillips. Global and local document degradation models. In *Proceedings of the Second International Conference on Document Analysis and Recognition*, pages 730–734, October 1993.
- [13] T. Kanungo, R.M. Haralick, H.S. Baird, W. Stuezel, and D. Madigan. A statistical, nonparametric methodology for document degradation model validation. *IEEE Transactions on PAMI*, 22(11):1209–1223, June 2000.
- [14] J. P. Lewis. Fast normalized cross-correlation. *Vision Interface*, 1995.
- [15] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing*, 41(12):3397–3415, 1993.
- [16] P. Marrgos and R.W. Schafer. Morphological filters, part 2: Their relations to median, order-statistic, and stack filters. *IEEE Transactions on acoustics, speech, and signal processing*, 35(8):87–134, 1987.
- [17] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, pages 40–44, 1993.
- [18] J.L. Starck, E.J. Candes, and D.L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on image processing*, 11(6):670–684, 2002.
- [19] V. N. Temlyakov. Weak greedy algorithms. *Advances in Computational Mathematics*, 5:173–187, 2000.
- [20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

At a Council held Nov. 9th 1753. [46 1778]
Upon considering the Petition of William Nelson
Esq. in behalf of himself and many others for a

Acc^t of the weather in Apr.
April 30. Very cold - wind
blowing exceeding hard
at N^o West at day. -
[faded text]

John Leasy
vs
Thomas Bowles
Affidavit of
Bowles

of government, is to do for
whatever they need to have
at all, or can not, so well
wth their separate, and in-

From
Capt. J. Whipple
his certificate in the
Collector's Office in
Providence N. H.
1769

(a)

At a Council held Nov. 9th 1753. [46 1778]
Upon considering the Petition of William Nelson
Esq. in behalf of himself and many others for a

Acc^t of the weather in Apr.
April 30. Very cold - wind
blowing exceeding hard
at N^o West at day. -
[faded text]

John Leasy
vs
Thomas Bowles
Affidavit of
Bowles

of government, is to do for
whatever they need to have
at all, or can not, so well
wth their separate, and in-

From
Capt. J. Whipple
his certificate in the
Collector's Office in
Providence N. H.
1769

(b)

At a Council held Nov. 9th 1753. [46 1778]
Upon considering the Petition of William Nelson
Esq. in behalf of himself and many others for a

Acc^t of the weather in Apr.
April 30. Very cold - wind
blowing exceeding hard
at N^o West at day. -
[faded text]

John Leasy
vs
Thomas Bowles
Affidavit of
Bowles

of government, is to do for
whatever they need to have
at all, or can not, so well
wth their separate, and in-

From
Capt. J. Whipple
his certificate in the
Collector's Office in
Providence N. H.
1769

(c)

Figure 7: (a) Noisy documents in DIBCO dataset used in Table 5, (b) Denoised documents got by our approach before binarization and (c) After binarization using Otsu's method.